

CT Foundation Model 기반 Attention Pooling을 활용한 복부 CT 다중 질환 분류 방법론

김강수^{*0}, 박재효[†], 배지영[‡], 정원기

고려대학교 정보대학 컴퓨터학과

{gangsusu1813, kjrnnny2, jy_bae, wkjeong}@korea.ac.kr

Multi-Disease Abdominal CT Classification Using Attention Pooling Based on CT Foundation Model Features

Gangsusu Kim^{*0}, Jaehyo Kwak[†], Jiyeong Bae[‡], Won-Ki Jeong

Department of Computer Science and Engineering, College of Informatics, Korea
University

{gangsusu1813, kjrnnny2, jy_bae, wkjeong}@korea.ac.kr

요 약

Computed Tomography(CT)는 복부 장기 질환 진단의 핵심 수단이나, 3D 볼륨 데이터의 고차원성과 양성 및 음성 클래스 불균형 문제는 자동 진단 모델 구축의 주요 장애물이다. 최근 대규모 사전학습 Foundation Model이 등장했으나, 이를 특정 질환 분류에 최적화하는 경량 어댑터 연구는 여전히 미비하다. 본 연구는 사전 학습된 CT Foundation Model의 3D 공간 특징을 효과적으로 활용하기 위해 Cross-Attention Pooling 기반의 경량 분류 방법론을 제안한다. 특히 관심 장기에 집중하는 Foreground mask 기법, 최적의 학습율을 탐색하는 Multi-LR Head 앙상블, 그리고 Logit Bias 기반의 임계 값 최적화를 통합하였다. 실험 결과, AMOS를 포함한 17개 질환 예측에서 평균 Balanced Accuracy 0.728을 기록하며 기존 베이스라인 대비 3.0%p의 유의미한 성능 향상을 입증하였다.

1. 서 론¹

Computed Tomography(CT)를 활용한 다중 질환 자동 진단은 임상 의사결정 효율화에 필수적이다. 최근 Medical Foundation Model(FM)의 등장으로 범용적인 3D 특징 추출이 가능해졌으나, 고차원 볼륨 데이터에서 특정 질환에 필요한 핵심 정보를 선별하는 어댑터(Adapter) 설계는 여전히 도전적인 과제이다. 특히 기존의 Linear Probe 방식은 단순히 평균을 내는 Global Average Pooling에 의존하여 세부 공간 정보를 소실하며, 극심한 데이터 불균형으로 인해 최적의 분류 임계값을 설정하는 데 한계가 있다.

본 연구에서는 이러한 한계를 극복하기 위해 Cross-Attention Pooling 기반의 경량 어댑터 구조를 제안한다. 첫째, 학습 가능한 Query 벡터를 활용해 질환별 중요 영역에 집중하는 Cross-Attention Pooling 구조를 도입하여 공간 정보 손실을

최소화했다. 둘째, 단일 학습 내에서 최적의 learning rate(lr)와 Feed Forward Network(FFN)의 용량을 결정하는 Multi-LR Head 및 Hyperparameter Sweep 전략을 통해 학습 효율을 극대화했다. 셋째, Logit Bias 최적화를 적용하여 클래스 불균형에 따른 결정 경계 편향 문제를 완화하였다. 실험 결과, AMOS 및 COVID-CT 등 17개 질환 타겟에서 기존 대비 +0.03 향상된 0.728의 Balanced Accuracy를 달성하였다. 본 연구는 최소한의 연산 자원으로 FM의 표현력을 극대화하여, 실제 의료 현장에 적합한 고성능, 경량형 진단 모델의 가능성을 제시한다.

2. 관련 연구

2.1 의료 영상 FM

최근 의료 영상 분야에서는 Swin UNETR[1]과 같은 Transformer 기반 모델과 VoCo[2] 등 자기 지도 학습을 통한 FM 연구[7]가 활발하다. 이러한 모델들은 레이블이 없는 CT 데이터로부터 범용적인 특징 표현을 추출하여 강력한 Backbone 역할을

[†] 공동 제 1 저자 (equal contribution)

수행하지만, 하위 task로의 효율적인 전이를 위한 경량 Adapter 설계가 필수적으로 요구된다.

2.2 Attention 기반 특징 추출

Transformer[8] 이후, Set Transformer[9]와 Attention-based MIL[10] 등 데이터의 중요 영역에 집중하는 메커니즘이 발전해 왔다. 특히 MIL 구조는 병리 영상과 같은 거대 이미지 분류에서 탁월한 성능을 입증하였다. 본 연구는 이를 확장하여 3D 특징에서 질환 관련 정보를 선택적으로 추출하는 Cross Attention Pooling 구조를 제안한다.

3. 문제 정의 및 데이터 셋

3.1 문제 정의

본 연구의 목표는 CT 볼륨으로부터 사전 학습된 FM 을 통해 추출된 3D 공간 특징 맵(Spatial Feature Map)을 입력으로 하여, 각 질환의 유무를 예측하는 이진 분류 모델을 구축하는 것이다. 입력 특징을 $F \in \mathbb{R}^{(D \times H \times W \times L)}$ ($D=1120, H=W=L=16$)이라 할 때, 모델은 각 질환에 대한 이진 레이블 $\{0,1\}$ 을 예측한다. 본 연구에서는 총 17개의 독립적인 질환에 대해 개별 분류기를 학습 및 평가한다. 모델의 성능 평가를 위해 클래스 불균형에 강건한 Balanced Accuracy (BACC)를 주 지표로, AUROC를 보조지표로 사용하며, 이는 다음과 같이 정의된다: $BACC = (Sensitivity + Specificity) / 2$

3.2 데이터셋 구성

본 연구에서는 복부 질환 진단의 범용성을 입증하기 위해 다음과 같은 3D CT 이미지 기반 대규모 데이터셋을 활용하였다.

3.2.1 AMOS [4] 15개의 복부 장기 및 이상 소견 분류를 수행한다. 특히 Total Segmentator 기반의 Foreground mask를 활용하여 고차원 특징 맵 내 배경 잡음을 억제하고, 실제 장기가 위치한 관심 영역(ROI)에 모델의 연산 자원을 집중시킨다.

3.2.2 LUNA25 [5] 폐 내부의 미세 결절 유무를 판별한다. 복부와 상이한 해부학적 환경 및 국소 병변 데이터를 통해 제안한 Cross-Attention Pooling의 정교한 특징 포착 능력과 도메인 전이 성능을 검증한다.

3.2.3 COVID-CT [6] 코로나19 감염에 따른 비정형적 폐 침윤 양상을 진단한다. 클래스 불균형이 극심한 특성을 활용하여, 본 연구의 Logit Bias 기반 임계 값 최적화 전략이 실제 데이터 분포에서 위양성/음성의 균형을 효과적으로 보정하는지 확인한다. 본 연구는 이처럼 상이한 특성을 가진 17개 타겟에 대한 통합적인 실험을 통해 제안 모델의 실용적 가치를 입증하였다.

4. 제안 방법론

4.1 전체 파이프라인

제안하는 다중 질환 진단 프레임워크는 (1) FM 기반 특징 추출, (2) Foreground Mask 를 활용한 ROI 특징 선택, (3) Cross-Attention Pooling 을 통한 정보 집약, (4) 질환별 최적 분류 및 학습 전략 적용의 4 단계로 구성된다.

4.2 Cross-Attention Pooling 기반 질환 어댑터

4.2.1 ROI 특징 선택 17개 질환 중 해부학적 위치가 명확한

11개 질환은 Foreground mask를 적용하여 해당 장기 영역의 특징만을 선택적으로 추출한다. 반면, 죽상동맥경화증, 대장암, 복수, 림프절병증과 같이 광범위하게 발생하는 4개 질환은 전체 CT Volume을 입력으로 사용한다.

4.2.2 Cross-Attention Pooling 추출된 특징 F_{ROI} 에서 질환별 핵심 정보를 집약하기 위해 학습 가능한 Query ($N_q = 2$)를 사용한다. 질환별 대표 표현 Z 는 다음과 같이 생성된다:

$$Z = LayerNorm \left(FFN \left(CrossATTN(Q, F_{ROI}, F_{ROI}) \right) \right)$$

4.2.3 분류 헤드 집약된 표현 Z 에 대해 Average Pooling을 수행한 후 선형 분류기를 거쳐 최종 클래스를 예측한다. 학습 시에는 CE Loss와 AdamW[11] 를 사용하며, 클래스 불균형 완화를 위해 Weighted Random Sampler를 적용한다.

4.3 Multi-LR 기반 앙상블 전략

질환마다 최적의 learning rate (lr)이 상이한 점을 고려하여, 동일한 Backbone 위에 9개의 분류 헤드를 병렬로 배치하였다. 각 헤드에 1×10^{-5} 부터 5×10^{-3} 사이의 서로 다른 lr 을 할당하여 최적의 학습 을 탐색한다. 최종 모델은 검증 데이터셋 에서 BACC 가 가장 높은 헤드를 선택한다.

4.4 Hyperparameter Sweep 및 학습 전략 다변화

학습 알고리즘의 최적화를 위해 다음과 같은 전략을 채택한다.

4.4.1 FFN 용량 탐색 FFN의 factor ratio ($ffn_{multti} \in \{1,2,4\}$)을 탐색하여 각 질환에 최적화된 파라미터를 결정한다.

4.4.2 Best-per-Target Selection 손실 함수와 스케줄러 조합에 따라 Recipe. A (Focal Loss [3] + Mixup [13] + Cosine Annealing [12])와 Recipe. B (CE Loss + Cosine Annealing [12]) 전략을 병렬 실험한 뒤, 각 질환별로 최고 성능을 보인 전략을 개별적으로 채택한다.

4.5 Logit Bias 기반 임계 값 최적화

추론 단계에서 클래스 불균형에 따른 편향을 보정하기 위해 Logit Bias b 를 도입한다. 양성 클래스의 출력에 Bias를 추가하여 결정 경계를 미세 조정한다. b 의 범위는 $[-3.0, +3.0]$ 구간에서 0.01 간격으로 탐색하며, 검증 데이터셋의 BACC 를 최대화하는 값을 최적의 임계 값으로 결정한다.

5. 실험

5.1 실험 설정

본 연구에서는 CTNexus 로부터 추출된 3D 특징을 입력으로 활용하였다. 추출에 사용된 FM의 가중치는 고정하였으며, 제안한 어댑터 구조에 대해서만 학습을 진행하였다.

5.2 단계별 성능 향상 분석

$\propto I$ 은 본 연구에서 제안한 각 전략이 모델 성능에 미치는 단계별 기여도를 나타낸다. Baseline(단일 lr 기반 Linear Probe) 대비 Multi-LR 9-Head 전략 도입 시 BACC가 0.014 향상되었으며, Hyperparameter Sweep과 전략 다변화 과정을 통해 0.027의 추가 향상을 기록하였다. 최종적으로 Logit Bias 기반 임계값 최적화를 적용했을 때 0.755의 BACC를 달성하며, 모든 제안 기법이 성능 개선에 유효함을 입증하였다. 구성

요소별 기여도를 분석하면, Multi-LR 및 Sweep을 통한 학습 전략 최적화가 +0.041 향상의 주된 요인이며, Logit Bias 기반 threshold 최적화는 +0.017로 학습 단계와 독립적으로 작동하는 추론 단계 보정 기법임을 확인하였다. Cross-Attention Pooling은 Global Average Pooling 대비 공간 정보를 보존하는 구조적 이점을 가지나, 단독 기여도의 정량적 분리는 향후 과제로 남긴다.

표 1. 제안 방법론의 단계별 성능 향상 (AMOS 기준)

	BACC	성능 향상
Baseline	0.697	-
+ Multi-LR	0.711	+0.014
+ Sweep & Recipe	0.738	+0.027
+ Logit Bias	0.755	+0.017

5.3 질환별 성능 및 한계점 분석

표 2는 17개 질환 타겟에 대한 Baseline과 제안 모델의 성능 비교 결과이다. 비장비대 (0.871), 수신증 (0.859), 지방간 (0.859) 등 형태적 변화가 뚜렷한 질환군에서 높은 성능을 보였다. 반면, 간낭종 (0.623)이나 림프절병증 (0.681) 등은 미세한 영상 소견과 양성 샘플 부족으로 상대적 낮은 성능을 기록하였다. 특히 LUNA25의 경우 도메인 차이와 미세 병변 특성상 성능 향상이 제한적이었으나, 복부 질환 전반에서는 성능 향상을 확인하였다. COVID-CT의 경우 사전학습 데이터와의 도메인 차이 및 폐 침윤 패턴 다양성으로 인해 baseline 대비 소폭 하락하였다. Liver Cyst는 양성 샘플 수가 부족하여 과적합이 발생한 것으로 분석된다.

표 2. 질환별 최종 성능 비교 (Base vs Ours)

질환	BACC		AUROC	
	Base	Ours	Base	Ours
Adrenal Hyperplasia	0.643	0.714	0.676	0.714
Ascites	0.691	0.750	0.691	0.764
Atherosclerosis	0.721	0.794	0.748	0.845
Cholecystitis	0.676	0.696	0.612	0.727
Colorectal Cancer	0.731	0.750	0.766	0.753
Fatty Liver	0.846	0.859	0.846	0.851
Gallstone	0.698	0.707	0.699	0.753
Hydronephrosis	0.813	0.859	0.896	0.865
Kidney Stone	0.658	0.684	0.693	0.715
Liver Calcifications	0.639	0.667	0.658	0.640
Liver Cyst	0.639	0.623	0.645	0.628
Liver Lesion	0.664	0.709	0.664	0.743
Lymphadenopathy	0.667	0.681	0.661	0.693
Renal Cyst	0.665	0.703	0.691	0.705
Splenomegaly	0.786	0.871	0.873	0.920
COVID-CT	0.766	0.750	0.779	0.739
LUNA25	0.575	0.570	0.561	0.491
Average	0.699	0.728	0.715	0.738

5.4 Hyperparameter Sweep의 효과

FFN factor 탐색 결과, AMOS 타겟 중 8개 에서 Sweep을 통해 결정된 모델이 최적 성능을 보였다. $ffn_{mult} = 1$ 이 최적인 경우가 4개, $ffn_{mult} = 2$ 와 4가 각각 2개 질환에서 최적으로 나타났다. 이는 질환의 복잡도와 발현 양상에 따라 요구되는 모델 구조의 상이함을 의미하며, 본 연구에서 채택한 질환별 최적 모델 선택 전략의 당위성을 뒷받침한다.

6. 결론

본 논문은 CT FM의 3D 공간 특징을 활용하여 복부 CT 다중 질환 이진 분류를 수행하는 파이프라인을 제안하였다. Cross-Attention Pooling 기반 경량 분류기에 Multi-LR Head 앙상블, Hyperparameter Sweep, Logit Bias Threshold 최적화를 결합하여 기준 성능 대비 0.03 향상된 0.728을 달성하였다. 단일 학습으로 9가지 lr 을 병렬 탐색하는 Multi-LR Head 방식으로 탐색 비용을 절감하였고, Target별 최고 선택 전략으로 질환별 최적 성능을 극대화하였으며, Logit Bias 최적화로 클래스 불균형 문제를 완화하였다. 향후 더 넓은 Hyperparameter 탐색, TTA 적용 및 낮은 성능을 보이는 질환에 특화 된 설계를 통한 추가 성능 향상을 기대한다.

참고 문헌

[1] A. Hatamizadeh et al., "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images." MICCAI Workshop, 2022.

[2] L. Wang et al., "Voco: A simple-yet-effective volume contrastive learning framework for 3d medical image analysis." CVPR, 2024.

[3] Lin, Tsung-Yi, et al. "Focal loss for dense object detection." ICCV, 2017.

[4] Ji, Yuanfeng, et al. "Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation." NeurIPS, 2022.

[5] D. Peeters et al., "The LUNA25 challenge: Public training and development set," 2025.

[6] D. Bell and T. Foster, "Covid-19. reference article." Radiopaedia, 2022.

[7] Tang, Yucheng, et al. "Self-supervised pre-training of swin transformers for 3d medical image analysis." CVPR, 2022.

[8] Vaswani, Ashish, et al. "Attention is all you need." NeurIPS, 2017.

[9] Lee, Juho, et al. "Set transformer: A framework for attention-based permutation-invariant neural networks." ICML, 2019.

[10] Ilse, Maximilian, Jakub Tomczak, and Max Welling. "Attention-based deep multiple instance learning." ICML, 2018.

[11] Loshchilov, Ilya, and Frank Hutter. "Decoupled weight decay regularization." ICLR, 2019.

[12] Loshchilov, Ilya, and Frank Hutter. "Sgdr: Stochastic gradient descent with warm restarts." ICLR, 2017.

[13] Zhang, Hongyi, et al. "mixup: Beyond empirical risk minimization." ICLR, 2018.